

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-173769

(43)Date of publication of application : 26.06.1998

(51)Int.Cl. H04M 1/65

 G10L 3/00

 G10L 3/00

 H04M 3/42

(21)Application number : 08-333309 (71)Applicant : MATSUSHITA ELECTRIC IND CO LTD

(22)Date of filing : 13.12.1996 (72)Inventor : TOKUDA TADAMICHI

(54) VOICE MESSAGE RETRIEVAL DEVICE

(57)Abstract:

PROBLEM TO BE SOLVED: To provide voice message retrieval that quickly retrieves contents of voice data.

SOLUTION: A voice message received by the retrieval device stores to a voice message storage section 3, while providing a time series label to the voice message, converts a keyword into a voice recognition model for an unspecified talker, when the user uses a key entry section 4 to enter the keyword desired by the user and sends the model to a voice recognition processing section 5, the voice-recognition processing section 5 uses the voice-recognition model for a voice message stored to apply voice-recognition processing of word spotting to the model and informs the voice message, including the keyword and its corresponding part to a control section 1. The control section 1 reproduces the keyword and parts by several seconds before and after the keyword, and the retrieval device reproduces the entire voice message by the designation of the user.

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平10-173769

(43)公開日 平成10年(1998)6月26日

| | | |
|--------------------------|-------|-------------|
| (51)Int.Cl. ⁶ | 識別記号 | F I |
| H 04 M 1/65 | | H 04 M 1/65 |
| G 10 L 3/00 | 5 3 1 | G 10 L 3/00 |
| | 5 5 1 | 5 3 1 C |
| H 04 M 3/42 | | 5 5 1 A |
| | | H 04 M 3/42 |
| | | J |

審査請求 未請求 請求項の数4 O.L (全 7 頁)

(21)出願番号 特願平8-333309

(71)出願人 000005821

松下電器産業株式会社

大阪府門真市大字門真1006番地

(22)出願日 平成8年(1996)12月13日

(72)発明者 徳田 肇道

大阪府門真市大字門真1006番地 松下電器
産業株式会社内

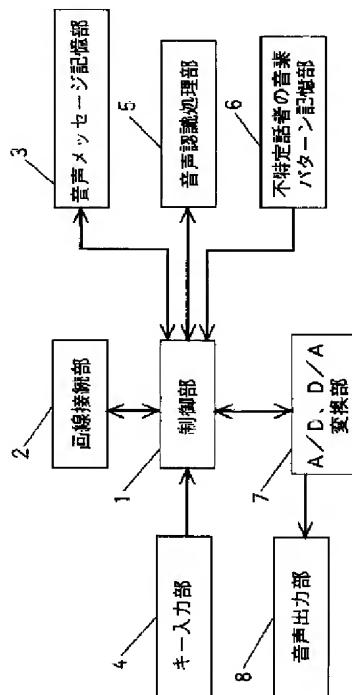
(74)代理人 弁理士 滝本 智之 (外1名)

(54)【発明の名称】 音声メッセージ検索装置

(57)【要約】

【課題】 音声メール蓄積装置などにおいて、蓄積された未分類の音声メッセージは内容検索ができないため、ユーザーは知りたい情報を得るためにそれらを順に全文再生する必要があり、その作業に非常に時間を要していた。

【解決手段】 装置が受け取った音声メッセージは時系列のラベルをつけて音声メッセージ記憶部3に蓄積し、ユーザーが検索したいキーワードをキー入力部4で入力すると、装置はそれを不特定話者の音声認識モデルに変換し、音声認識処理部5に送り、音声認識処理部5では蓄積された音声メッセージを対象に音声認識モデルを用いてワードスポットティングの音声認識処理を行い、キーワードを含む音声メッセージとその該当部分を制御部1に通知し、制御部1はキーワードとその前後数秒の部分を再生し、さらにユーザーの指定により音声メッセージの全文を再生する音声メッセージ検索装置とする。



【特許請求の範囲】

【請求項1】音声によるメッセージを蓄積する手段と、検索手段を有し、全体の処理を制御する制御部と、外部の電話回線、構内回線と装置を接続する回線接続部と、ユーザーが音声メッセージの検索キーワードを入力するためのキー入力部と、キーワードの文字列を不特定話者の音素パターンの連鎖に変換するために参照する不特定話者の音素パターン記憶部と、音声メッセージに対してキーワードスポットティング音声認識を行う音声認識処理部と、音声信号のA/D変換およびD/A変換処理を行うA/D・D/A変換部と、検索結果を音声として再生する音声出力部とを備えたことを特徴とする音声メッセージ検索装置。

【請求項2】ユーザーが検索キーワードを音声で入力するための音声入力部と、発声されたキーワードを特定話者の音素列に変換するために参照される特定話者の音素パターン記憶部を加え備えた請求項1記載の音声メッセージ検索装置。

【請求項3】検索キーワードの音声を入力するための音声入力部と、不特定多数のユーザーに発声されたキーワードを単語認識処理により単語標準パターンに変換するために参照される不特定話者の単語標準パターン記憶部を加え備えた請求項1記載の音声メッセージ検索装置。

【請求項4】検索キーワードの音声を入力するための音声入力部と、不特定多数のユーザーが一音ずつ区切って発声したキーワードを単音節認識処理するために参照される不特定話者の単音節標準パターン記憶部を加え備えた請求項1記載の音声メッセージ検索装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、音声メールの蓄積装置や、留守番電話機などにおいて、受信した音声メッセージを蓄積し、ユーザーの指定によりそれらを検索する機能を有する音声メッセージ検索装置に関する。

【0002】

【従来の技術】一般に音声メールの蓄積装置や留守番電話機など、音声によるメッセージを蓄積・再生する装置では、受け取ったメッセージを時間順に記憶し、ユーザーの指示で順番に再生する機能を有している。また、ユーザーはメッセージに手動でラベルを付加してキーワード別に分類することにより、それらを検索することができる。

【0003】

【発明が解決しようとする課題】ところで、前記の音声によるメッセージを蓄積・再生する装置では、装置に蓄積された未分類の音声メッセージは内容検索ができないため、ユーザーは知りたい情報を得るためにそれらを順に全文再生する必要があり、その作業はメッセージ件数が多いほど非常に時間がかかっていた。また、特定のキーワードを発声内容として含む音声メッセージが分類さ

れ、その内容を再生して確認する際、キーワードの周辺を聞くだけで内容が分かる場合でもメッセージの全文が再生されるため、時間が余分にかかっていた。

【0004】本発明は前記従来の問題に留意し、音声データの迅速な内容検索を可能とする音声メッセージ検索装置を提供することを目的とする。

【0005】

【課題を解決するための手段】前記目的を達成するため、本発明は、ユーザーが音声メッセージの検索キーワードを入力するためのキー入力部と、キーワードの文字列を不特定話者の音素パターンの連鎖に変換するために参照する不特定話者の音素パターン記憶部と、音声メッセージに対してキーワードスポットティング音声認識を行う音声認識処理部と、音声信号のA/D変換およびD/A変換処理を行うA/D・D/A変換部と、検索結果を音声として再生する音声出力部とを備えた音声メッセージ検索装置としたものである。

【0006】本発明によれば、音声データの迅速な内容検索を可能とする音声メッセージ検索装置とができる。

【0007】

【発明の実施の形態】本発明の請求項1に記載の発明は、音声によるメッセージを蓄積する手段と、検索手段を有し、全体の処理を制御する制御部と、外部の電話回線、構内回線と装置を接続する回線接続部と、ユーザーが音声メッセージの検索キーワードを入力するためのキー入力部と、キーワードの文字列を不特定話者の音素パターンの連鎖に変換するために参照する不特定話者の音素パターン記憶部と、音声メッセージに対してキーワードスポットティング音声認識を行う音声認識処理部と、音声信号のA/D変換およびD/A変換処理を行うA/D・D/A変換部と、検索結果を音声として再生する音声出力部とを備えた音声メッセージ検索装置であり、発声内容にキーワードを含む音声メッセージを検索することを可能とし、また、メッセージの全文を聞かなくても内容の確認を可能にし、確認に要する時間を短縮させるという作用を有する。

【0008】なお、ここでワードスポットティングとは、音声パターンに対して区間を限定せずに標準パターンとの参照を行い、一致尤度が高い部分を探すことにより、目的の単語を検出する認識方式である。

【0009】本発明の請求項2に記載の発明は、請求項1記載の音声メッセージ検索装置に、ユーザーが検索キーワードを音声で入力するための音声入力部と、発声されたキーワードを特定話者の音素列に変換するために参照される特定話者の音素パターン記憶部を加え備えた音声メッセージ検索装置であり、キーワードをユーザーが発声して音声メッセージを検索でき、特定話者の音素認識により、検索キーワードを高い精度で音声入力することができるという作用を有する。

【0010】本発明の請求項3に記載の発明は、請求項1記載の音声メッセージ検索装置に、検索キーワードの音声を入力するための音声入力部と、不特定多数のユーザーに発声されたキーワードを単語認識処理により単語標準パターンに変換するために参照される不特定話者の単語標準パターン記憶部を加え備えた音声メッセージ検索装置であり、検索キーワードの入力において、キーワードを不特定多数のユーザーが発声し、それを不特定話者の単語標準パターンを用いて認識することにより、特定のユーザーに限らず音声検索機能を利用することを可能とするという作用を有する。

【0011】本発明の請求項4に記載の発明は、請求項1記載の音声メッセージ検索装置に、検索キーワードの音声を入力するための音声入力部と、不特定多数のユーザーが一音ずつ区切って発声したキーワードを単音節認識処理するために参照される不特定話者の単音節標準パターン記憶部を加え備えた音声メッセージ検索装置であり、検索キーワードの入力において、キーワードを不特定多数のユーザーが発声し、それを不特定話者の単語標準パターンを用いて認識するので、特定のユーザーに限らず音声検索機能を利用することを可能とする作用を有する。

【0012】以下、本発明の音声メッセージ検索装置の実施の形態を図面を参照して説明する。

【0013】(実施の形態1) 図1は本発明の実施の形態1の音声メッセージ検索装置の構成を示すブロック図、図2は同音声メッセージ検索装置における音声認識装置の回路ブロック図、図3は同音声メッセージ検索装置の動作を示すフローチャートである。

【0014】図1の音声メッセージ検索装置の構成ブロック図において、構成要素として1は制御部、2は回線接続部、3は音声メッセージ記憶部、4はキー入力部、5は音声認識処理部、6は不特定話者の音素パターン記憶部、7はA/D・D/A変換部、8は音声出力部である。

【0015】また、図2の音声認識装置の回路ブロック図において、構成要素として21はキーボード、22はスピーカー、23はマイク、24は中央処理装置(CPU)、25は書き込み可能メモリ(RAM)、26は読みだし専用メモリ(ROM)である。

【0016】前記構成要素において、キー入力部4はキーボード21により、音声入力部9はマイク23により、音声出力部8はスピーカー22により、音声メッセージ記憶部3はRAM25により、不特定話者の音素パターン記憶部6はROM26によりそれぞれ実行され、また、制御部1と音声認識処理部5とA/D・D/A変換部7はCPU24がROM26に書かれたプログラムを実行し、RAM25にアクセスすることにより実行される。

【0017】上記の各構成要素よりなる音声メッセージ

検索装置について、その動作を図3のフローチャートに従い説明する。なお、装置が受け取った音声メッセージは、時系列順にラベルをつけて音声メッセージ記憶部3に格納されているものとする。

【0018】step1では、ユーザーが検索したいキーワードをキーボード21で入力する。入力は直接、あるいはネットワークを経由して行われる。

【0019】step2では、不特定話者の音素パターン記憶部6に格納されている不特定話者の音素パターンを参照してキーワードの文字列を不特定話者の音声認識パターンに規則合成する。

【0020】step3では、音声認識処理部5は合成された音声認識パターンを受け取り、蓄積されている音声メッセージを音声メッセージ記憶部3からロードして順にワードスボッティングの音声処理を行う。認識尤度は結果判定部に送られ、メッセージ中でキーワードを抽出したと判断された箇所がメモリーに記録される。

【0021】step4では、メモリーを参照し、キーワードが検出されたメッセージの該当箇所がヒット順に音声としてスピーカー22から再生される。または、ネットワーク経由で音声データが送られ、ユーザーの端末に接続されたスピーカーから再生される。再生範囲はキーワードを含む前後数秒、または、音声パワーの計算により発声の区切りを疑似的に求めた文章単位とする。

【0022】step5では、ユーザーが再生された音声を聞いてメッセージの全文を聞くか否かを判断する。したがって、一部を聞くだけで内容が分かった場合は全文を聞くことなく済むことができる。また、キーワードを含まないメッセージが誤認識によってヒットした場合も、その部分を聞くことにより事前に判定できるので、目的と異なるメッセージが全文再生されることを避けることができる。

【0023】step6では、ユーザーが指定した場合のみ、メッセージの全文を再生する。

【0024】step7では、検索対象の全メッセージについて、step3～step6を繰り返す。

【0025】このように、この実施の形態1では、ユーザーが検索キーワードをキーボードより入力し、その文字列を不特定話者の音素標準パターンの連鎖に変換し、蓄積された音声メッセージに対して不特定話者のワードスボッティングの音声認識処理を行うことにより、発声内容にキーワードを含む音声メッセージを検索することを可能とし、また、メッセージ中でキーワードが検出された場合、その前後数秒、あるいは音声パワーにより求めた文章単位のみを再生することにより、メッセージの全文を聞かなくても内容の確認を可能にし、確認に要する時間を短縮させることができる。

【0026】(実施の形態2) 図4は本発明の実施の形態2の音声メッセージ検索装置の構成を示すブロック図、図5は同音声メッセージ検索装置の動作を示すフロ

ーチャートである。

【0027】図4の音声メッセージ検索装置の構成ブロック図において、符号1から8までの各構成要素は前述の実施の形態1と同一であり、かつ、同様に構成されている。

【0028】この実施の形態2の特徴は、前述の実施の形態1の構成に、音声入力部9と、特定話者の音素パターン記憶部10を加え備えたことにある。

【0029】なお、前述の図2の音声メッセージ検索装置における音声認識装置の回路ブロック図は説明に共用するが、ここではROM26が特定話者の音素パターン記憶部10を実行する。

【0030】上記の各構成要素よりなる音声メッセージ検索装置について、その動作を図5のフローチャートに従い説明する。この実施の形態2の音声メッセージ検索装置では、前述の実施の形態1のstep1の代わりに次の処理を行う。

【0031】step1では、ユーザーが検索したいキーワードを音声入力手段に発声する。発声は装置に直接、あるいは外部から電話回線を通してなされる。音声信号はA/D・D/A変換部7によりデジタル信号に変換される。

【0032】step2では、音声信号を音声認識処理部5に転送し、特定話者の音素パターン記憶部10を参照して特定話者の音素認識処理を行う。音素パターンは前処理としてユーザーの発声を音響的セグメンテーションすることにより得られる。

【0033】音響的セグメンテーションの方法としては、発声内容が既知の音声データに対して、次式で定義されるフレーム時刻i,j間のセグメント内距離D(i,j)の総和のフレーム平均が予め定められた歪みより小さくなるようセグメント境界を求める。

【0034】

【数1】

$$D(i,j) = \sum_{m=i}^{j-1} (x_m - x'_m)^2$$

【0035】ここで、 x_m は特徴ベクトル、 x'_m はフレーム時刻iからjの間の平均軌跡ベクトルである。求められたセグメントの平均ベクトルと分散は発声内容の音素配列と対応して記録される。対応関係が不明な場合は過去に蓄積された音素データを用いてダイナミックプログラミングにより推定し、補間していく。

【0036】特定話者の音素パターンを得る別の方法としては、発声内容が既知の音声内容が既知の音声データに対して、不特定話者の音素パターンを用いてマッチングを行い、各音素に対応する区間の特徴ベクトルを教師信号として適応化処理を行う方法がある。また、教師信号を得なくても、対象話者の音声の認識とその認識結果を教師とした話者適応を交互に繰り返すことにより、音素パターンを徐々に話者に適応化させる方法も考えられ

る。

【0037】step3では、認識結果の音素列を音声信号に変換して再生し、認識結果が正しいことをユーザーが確認した上でキーワードの文字列を制御部に渡す。

【0038】このように実施の形態2では、検索キーワードの入力において、キーワードをユーザーが発声し、それを特定話者の音素パターンを用いて音声認識処理を行い、さらに認識結果を不特定話者の音素標準パターンの連鎖に変換し、それを用いて音声メッセージを検索する。特定話者の音素認識により、検索キーワードを高い精度で音声入力することができる。また、任意のキーワードを用いることができるという作用を有する。

【0039】なお、特定話者の音素パターンは、ユーザーが発声した音声をA/D変換し、自動セグメンテーション（音素のラベル付け）して音素パターンを切り出すか、不特定話者の音素パターンをユーザーの音声特徴に適応化することにより作成しておく。

【0040】（実施の形態3）図6は本発明の実施の形態3の音声メッセージ検索装置の構成を示すブロック図、図7は同音声メッセージ検索装置の動作を示すフローチャートである。

【0041】図6の音声メッセージ検索装置の構成ブロック図において、符号1から8までの各構成要素は前述の実施の形態1と同一であり、かつ、同様に構成されている。

【0042】この実施の形態3の特徴は、前述の実施の形態1の構成に、音声入力部9と、不特定話者の単語音声パターン記憶部11を加え備えたことにある。

【0043】なお、前述の図2の音声メッセージ検索装置における音声認識装置の回路ブロック図は説明に共用するが、ここではROM26が不特定話者の単語音声パターン記憶部11を実行する。

【0044】上記の各構成要素よりなる音声メッセージ検索装置について、その動作を図7のフローチャートに従い説明する。この実施の形態3の音声メッセージ検索装置では、前述の実施の形態1のstep1およびstep2の代わりに次の処理を行う。

【0045】step1では、ユーザーが検索したいキーワードを音声入力手段に発声する。発声は装置に直接、あるいは外部から電話回線を通してなされる。音声信号はA/D・D/A変換部7によりデジタル信号に変換される。

【0046】step1およびstep2では、音声信号を音声認識処理部5に転送し、不特定話者の単語音声パターン記憶部11を参照して不特定話者の単語認識処理を行う。認識単位は単語、および単語の連結とする。たとえば、「企画」と「会議」の単語音声パターンがある場合、「企画会議」もキーワードとして認識できる。

【0047】step3では、認識結果を音声信号で再生し、結果が正しいか否かをユーザーが確認する。

【0048】step 4では、キーワードとして認識された不特定話者の単語音声パターンをそのまま制御部1に渡し、音声メッセージの検索に用いる。

【0049】キーワード入力時の認識処理および音声メッセージのワードスポットティング認識処理の精度を音素パターンの場合よりも高くすることができる。

【0050】このように実施の形態3では、検索キーワードの入力において、キーワードを不特定多数のユーザーが発声し、それを不特定話者の単語標準パターンを用いて認識することにより、特定のユーザーに限らず音声検索機能を利用することを可能とする。単語標準パターンは単独、またはそれらの連鎖で用いることにより、認識可能な語彙を増やす。認識されたキーワードの単語標準パターンは、そのまま音声メッセージのワードスポットティング認識処理に用いることができる。単語パターンの使用により、キーワード入力時の認識処理および音声メッセージのワードスポットティング認識処理の精度を音素パターンの場合よりも高くすることができる。

【0051】(実施の形態4) 図8は本発明の実施の形態4の音声メッセージ検索装置の構成を示すブロック図、図9は同音声メッセージ検索装置の動作を示すフローチャートである。

【0052】図8の音声メッセージ検索装置の構成ブロック図において、符号1から8までの各構成要素は前述の実施の形態1と同一であり、かつ、同様に構成されている。

【0053】この実施の形態4の特徴は、前述の実施の形態1の構成に、音声入力部9と、不特定話者の単音節標準パターン記憶部12を加え備えたことにある。

【0054】なお、前述の図2の音声メッセージ検索装置における音声認識装置の回路ブロック図は説明に共用するが、ここではROM26が不特定話者の単音節標準パターン記憶部12を実行する。

【0055】上記の各構成要素よりなる音声メッセージ検索装置について、その動作を図7のフローチャートに従い説明する。この実施の形態4の音声メッセージ検索装置では、前述の実施の形態1のstep1およびstep2の代わりに次の処理を行う。

【0056】step1では、ユーザーが検索したいキーワードを音節ごとに区切って発声する。発声は装置に直接、あるいは外部から電話回線を通してなされる。音声信号はA/D・D/A変換部7によりデジタル信号に変換される。

【0057】step2では、音声信号を音声認識処理部5に転送し、不特定話者の単音節標準パターン記憶部12を参照して単音節の認識処理を行う。

【0058】step3では、認識結果の単音節をその都度、音声信号で再生し、結果が正しいか否かをユーザーが確認する。

【0059】step4では、キーワードとして認識さ

れた不特定話者の単音節パターンをそのまま連結して単語標準パターンを作り、制御部1に渡し、音声メッセージの検索に用いる。

【0060】このように実施の形態4では、検索キーワードの入力において、キーワードを不特定多数のユーザーが発声し、それを不特定話者の単語標準パターンを用いて認識することにより、特定のユーザーに限らず音声検索機能を利用することを可能とする。認識されたキーワードの文字列は一定の規則により不特定話者の音素パターンの連鎖に合成され、音声メッセージのワードスポットティング認識処理に用いられる。不特定話者の単音節標準パターンの使用により、特定のユーザーに限定されず、しかも任意の言葉を検索キーワードに指定することができる。

【0061】

【発明の効果】前記の説明より明らかなように本発明は、未分類の音声メッセージに対してキー入力によるキーワードで検索を行い、検索結果を効率良く確認することができる。標準パターンとして音素の連鎖を用いるため、部分一致、完全一致などの条件を指定できる。また、予めキーワードを指定し、メッセージ受信時に自動的に分類保存することも可能である。

【0062】また、メッセージ中でキーワード付近のみを再生することにより、多くの場合は全文を聞かなくても内容が確認できるため、ヒットしたメッセージが多いときに再生時間を大幅に節約することができる。そして、誤認識でヒットした場合には、メッセージ全文を聞く前にキャンセルすることができる。

【0063】また、検索キーワードの音声入力が可能になる。たとえば、ユーザーは外出先から電話をかけて音声メッセージの検索を行うこともできる。また、キーワードの認識に特定話者の音素パターンを用い、キーワードの語彙に特定の制限がなく、かつ、認識精度の低下を抑えることができる。

【0064】また、キーワードの認識に、不特定話者の単語単位の標準パターンあるいは単音節単位の標準パターンを使用することにより、特定のユーザーに限らず利用することができる。

【図面の簡単な説明】

【図1】本発明の実施の形態1の音声メッセージ検索装置の構成を示すブロック図

【図2】同音声メッセージ検索装置における音声認識装置の回路ブロック図

【図3】同音声メッセージ検索装置の動作を示すフローチャート

【図4】本発明の実施の形態2の音声メッセージ検索装置の構成を示すブロック図

【図5】同音声メッセージ検索装置の動作を示すフローチャート

【図6】本発明の実施の形態3の音声メッセージ検索装置

置の構成を示すブロック図

【図7】同音声メッセージ検索装置の動作を示すフローチャート

【図8】本発明の実施の形態4の音声メッセージ検索装置の構成を示すブロック図

【図9】同音声メッセージ検索装置の動作を示すフローチャート

【符号の説明】

- 1 制御部
- 2 回線接続部
- 3 音声メッセージ記憶部
- 4 キー入力部
- 5 音声認識処理部
- 6 不特定話者の音素パターン記憶部
- 7 A/D・D/A変換部
- 8 音声出力部

6 不特定話者の音素パターン記憶部

7 A/D・D/A変換部

8 音声出力部

9 音声入力部

10 特定話者の音素パターン記憶部

11 不特定話者の単語音声パターン記憶部

12 不特定話者の単音節標準パターン記憶部

21 キーボード

22 スピーカー

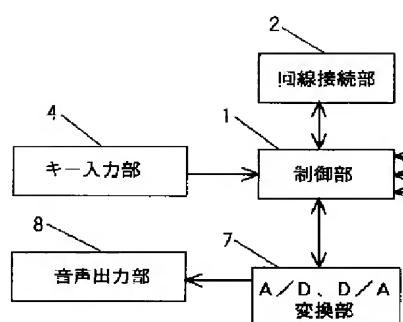
23 マイク

24 CPU

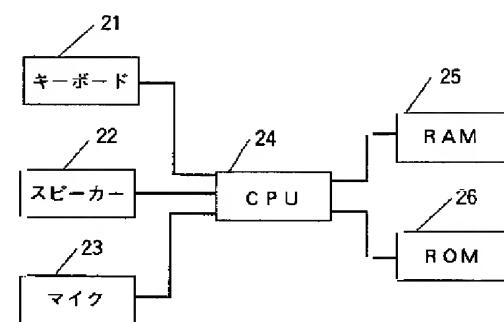
25 RAM

26 ROM

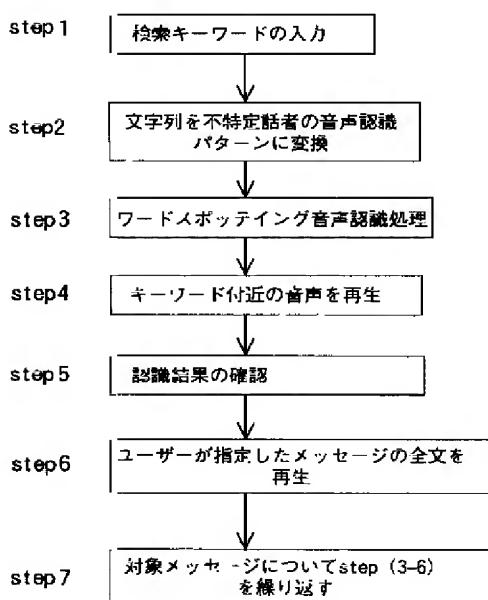
【図1】



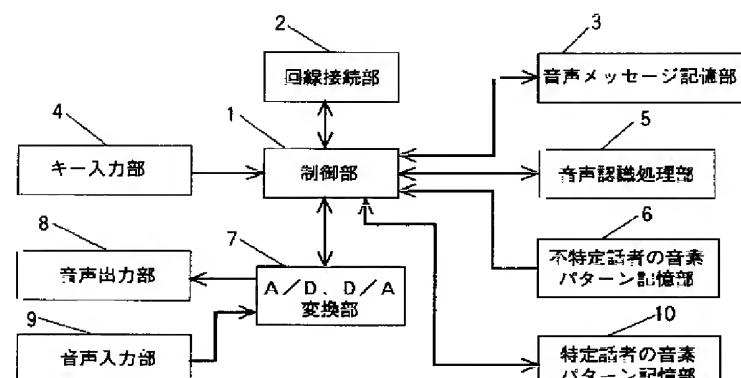
【図2】



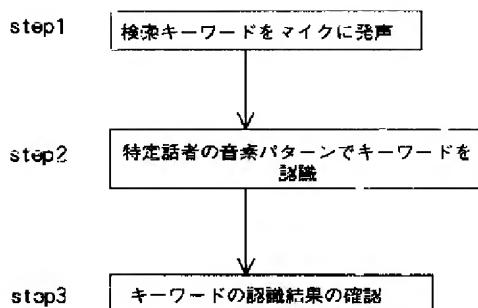
【図3】



【図4】

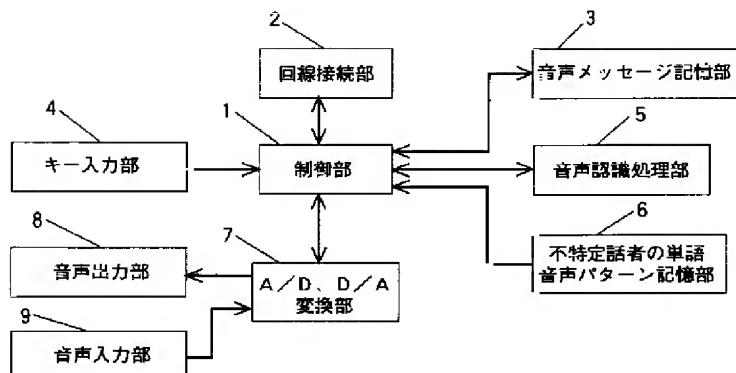


[図5]



【図7】

【圖6】



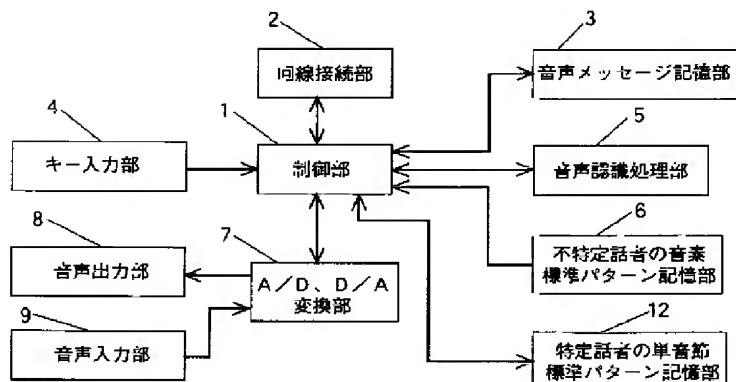
```

graph TD
    A[検索キーワードをマイクに発声] --> B[不特定話者の単語音声パターンでキーワードを認識]
    B --> C[キーワードの認識結果の確認]
    C --> D[不特定話者の単語音声パターンを制御部に渡す]

```

〔四九〕

[図8]



```

graph TD
    A[検索キーワードをマイクに区切って発声] --> B[不特定話者の単音節標準パターンで  
キーワードを認識]
    B --> C[キーワードの認識結果の確認]
    C --> D[不特定話者の単語音声パターンを合成し、  
制御部に渡す]
    
```

The flowchart illustrates a four-step process for keyword recognition. Step 1 involves dividing search keywords into single音节 segments and pronouncing them into a microphone. Step 2 involves recognizing these keywords using standard patterns for non-native speakers. Step 3 involves confirming the recognition results. Step 4 involves synthesizing the single word sound patterns and passing them to the control unit.